CLAIMS

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

1. A method of executing a linear algebra subroutine on a computer having at least one cache, said method comprising:

streaming data for matrices involved in processing said linear algebra subroutine such that data is processed using data for a first matrix residing in said cache as one of an entirety of said first matrix and a submatrix of said first matrix and data from a second matrix and a third matrix is respectively residing as one of its entirety and submatrices thereof in a memory device at a higher level than said cache,

said streaming providing data from said higher level as said data is required for said processing.

2. The method of claim 1, wherein said at least one cache comprises an L1 cache and said higher level comprises an L2 cache.

3. The method of claim 1, further comprising:

selecting said matrix stored in said cache by determining which matrix will fit into said cache.

4. The method of claim 3, further comprising:

determining a size of each of said first matrix, said second matrix, and said third matrix;

determining which of said first matrix, said second matrix, and said third matrix will fit into a size of said cache; and

loading data for a selected one of said first matrix, said second matrix, and said third matrix into said cache.

5. The method of claim 1, further comprising:

selecting a linear algebra subroutine from a plurality of subroutines to perform a matrix operation, said selecting based on which of said plurality of subroutine has a format consistent with said matrix stored in said cache.

6. The method of claim 2, wherein data for said second matrix and said third matrix streams into said L1 cache from said L2 cache such that said data from one of said second matrix and said third matrix streams in a vector format and data from the other of said second matrix and said third matrix streams in a scalar format.

7. The method of claim 1, wherein said linear algebra subroutine comprises a subroutine from a LAPACK (Linear Algebra PACKage).

8. The method of claim 7, wherein said LAPACK subroutine comprises a BLAS Level 3 L1 cache kernel.

9. An apparatus, comprising:

a memory system to store matrix data for processing in a linear algebra program using data from a first matrix, a second matrix, and a third matrix, said memory system including at least one cache; and

a processor to perform a linear algebra operation, wherein data from one of said first matrix, said second matrix, and said third matrix is stored in said cache in a matrix format and data from a remaining two matrices is stored in said memory system at a level higher that said cache,

said data from said remaining two matrices being streamed into said processor as required by said processing.

10. The apparatus of claim 9, further comprising:

a selector to determine a size of each of matrices involved in a matrix multiplication process and to select one of said matrices to reside in said cache, as based on having determined said sizes;

a loader to load data for the selected matrix into said cache; and

a selector to select a matrix subroutine to perform said linear algebra program processing,

said selected matrix subroutine having a format consistent with which said matrix is selected to reside in said cache.

11. The apparatus of claim 9, wherein said linear algebra program comprises a subroutine from a LAPACK (Linear Algebra PACKage).

12. The apparatus of claim 11, wherein said LAPACK subroutine comprises a BLAS Level 3 L1 cache kernel types.

13. The apparatus of claim 10, wherein said plurality of matrix subroutines comprises two of six possible matrix subroutines.

14. A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method of

of executing a linear algebra subroutine on a computer having at least one cache, said method comprising:

streaming data for up to three matrices involved in processing said linear algebra subroutine such that data is processed using data for a first matrix stored in said cache as a matrix format and data from a second matrix and a third matrix is stored in a memory device at a higher level than said cache, said streaming providing data from said higher level in a manner as said data is required for said processing.

15. The signal-bearing medium of claim 14, said method further comprising:

determining a size of each of matrices involved in a matrix multiplication process;

selecting one of said matrices to reside in one of said at least one cache, based on having determined said sizes; and

selecting a matrix subroutine from a plurality of subroutines by determining which said matrix subroutine can perform said matrix multiplication consistent with which matrix is selected to reside in said one of said at least one cache.

16. The signal-bearing medium of claim 14, wherein said matrix subroutine comprises a subroutine from LAPACK (Linear Algebra PACKage).

17. The signal-bearing medium of claim 16, wherein said LAPACK subroutine comprises a BLAS Level 3 L1 cache kernel types.

18. The signal-bearing medium of claim 14, wherein data for said second matrix and said third matrix streams from said higher level such that said data from one of said second matrix and said third matrix streams in a vector format and data from the other of said second matrix and said third matrix streams in a scalar format.

19. A method of providing a service involving at least one of solving and applying a scientific/engineering problem, said method comprising at least one of:

using a linear algebra software package that performs one or more matrix processing operations, said method comprising streaming data for matrices involved in processing said linear algebra subroutines such that data is processed using data for a first matrix stored in a cache as a matrix format and data from a second matrix and a third matrix is stored in a memory device at a higher level than said cache, said streaming providing data from said higher level in a manner as said data is required for said processing;

providing a consultation for solving a scientific/engineering problem using said linear algebra software package;

transmitting a result of said linear algebra software package on at least one of a network, a signal-bearing medium containing machine-readable data representing said result, and a printed version representing said result; and

receiving a result of said linear algebra software package on at least one of a network, a signal-bearing medium containing machine-readable data representing said result, and a printed version representing said result.

20. The method of claim 19, wherein said matrix subroutines comprise BLAS Level 3 L1 cache kernels from a LAPACK (Linear Algebra PACKage).